A Novel Hybrid Model to Predict Dissolved Oxygen for Efficient Water Quality in Intensive Aquaculture

Chippada Pavan

Reg. No. 24Q71F0008

Pavanchippada5@gmail.com

Department of Master of Computer Applications

Avanthi Institute of Engineering and Technology (Autonomous)

Vizianagaram, Andhra Pradesh, India

Under the guidance of Mr. S.VASANTH, MCA, Assistant Professor

vasanthsunkara421@gmail.com

Abstract—*The accurate prediction of dissolved oxygen (DO) levels is critical for maintaining water quality in intensive aquaculture systems, where the health and growth of aquatic species are heavily dependent on oxygen availability. This study presents a novel hybrid model designed to enhance the prediction accuracy of dissolved oxygen, leveraging the strengths of both machine-learning algorithms and traditional physical models. The hybrid model integrates data-driven approaches such as Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, and Support Vector Machines (SVM) with numerical simulations, allowing it to capture both the non-linear patterns of DO fluctuations and the underlying environmental factors. The model architecture is developed using real-time sensor data collected from aquaculture systems, including variables such as water temperature, pH, salinity, and nutrient concentrations, which are processed through the hybrid model to forecast dissolved oxygen levels with improved precision over conventional methods. By incorporating explainable AI techniques, the model provides insight into the contribution of each variable, enhancing interpretability for aquaculture operators and improving decision-making regarding oxygen-regulation practices. In validation tests using historical and real-time data, the novel hybrid model demonstrated superior performance compared with standalone machine-learning models, achieving higher accuracy and lower error rates in DO prediction. This advancement offers a reliable tool for farmers to ensure optimal oxygen levels, reduce fish mortality, and improve overall aquaculture productivity, paving the way for smarter, more sustainable water-management solutions.*

Keywords—*Dissolved Oxygen Prediction; Intensive Aquaculture; Hybrid Model; LSTM; Support Vector Machine; Artificial Neural Network; Explainable AI; Water Quality.*

I. INTRODUCTION

In the rapidly expanding field of aquaculture, maintaining optimal water quality is essential for the sustainability and productivity of intensive fish-farming operations. One of the most critical factors affecting water quality is the concentration of dissolved oxygen (DO), which directly influences the health, growth, and survival of aquatic organisms. Inadequate levels of DO can lead to stress, disease, and mortality in fish, while excessive oxygen can increase operational costs. Therefore, precise monitoring and control of DO are vital for ensuring both the efficiency of aquaculture systems and the well-being of cultured species.

Traditional methods of predicting dissolved oxygen levels rely heavily on empirical models and fixed equations, which often fall short in capturing the complex, dynamic interactions between environmental variables such as temperature, salinity, and biological oxygen demand. These conventional approaches can struggle with the non-linear and time-variant nature of DO fluctuations, leading to inaccuracies that could compromise water-management decisions. In recent years, the growing availability of sensor-based monitoring systems has generated large volumes of real-time data, opening up new opportunities for advanced prediction models that can adapt to changing environmental conditions.

This work proposes a novel hybrid model for predicting dissolved oxygen levels in aquaculture systems, combining the strengths of machine-learning techniques with traditional numerical models. The hybrid approach leverages the power of artificial neural networks to capture non-linear relationships within the data, while the numerical models incorporate domain-specific knowledge of oxygen dynamics and water chemistry. In addition to improving prediction accuracy, the hybrid model addresses interpretability through explainable AI (XAI) techniques, enabling users to understand how input variables such as water temperature, pH, and nutrient levels impact DO predictions, thereby supporting informed adjustments to oxygen levels and other critical parameters for more sustainable and profitable aquaculture practices.

II. LITERATURE SURVEY

The prediction of dissolved oxygen and other water-quality parameters in aquaculture has been widely studied, with a clear trend toward hybrid models that combine multiple machine-learning techniques. Wang, Xu, and Chen presented a hybrid model combining neural networks and support vector machines to predict dissolved oxygen levels in aquaculture environments, integrating historical DO data with real-time sensor inputs to enhance prediction accuracy and highlighting the benefits of combining different machine-learning techniques to capture complex patterns in water-quality data. Liu, Zhang, and Yang introduced an integrated hybrid approach merging artificial neural networks with fuzzy-logic systems to predict multiple water-quality parameters, including dissolved oxygen, by leveraging both historical data and real-time measurements.

Broader research reinforces these directions. Studies on neural-network-based DO prediction emphasise the advantages of advanced computational methods over empirical equations, while reviews of data-driven approaches highlight the role of machine learning in water-quality management. Work on explainable AI in environmental science underscores the importance of interpretability for complex prediction models, and studies of ensemble methods and real-time monitoring demonstrate improvements in accuracy and responsiveness. Collectively, the literature shows that hybrid models integrating temporal deep-learning methods such as LSTM with robust classifiers such as SVM—and incorporating explainability—offer superior performance for DO prediction, which motivates the proposed system.

TABLE I. SUMMARY OF REPRESENTATIVE PRIOR WORK

S.No	Author(s) / Year	Methodology	Contribution
1	Wang, Xu & Chen	NN + SVM hybrid	DO prediction from historical + sensor data
2	Liu, Zhang & Yang	ANN + fuzzy logic	Multi-parameter water-quality prediction
3	Lee & Kim, 2019	Neural networks	DO prediction in aquaculture
4	Garcia & Torres, 2020	Explainable AI	Interpretability in environmental science
5	Singh & Patel, 2023	Ensemble methods	Predictive modelling of aquatic ecosystems
6	Yang & Zhao, 2021	Real-time monitoring	Temporal dynamics of DO

III. EXISTING SYSTEM AND PROPOSED SYSTEM

A. Existing System

The existing systems for predicting dissolved oxygen levels in intensive aquaculture primarily rely on traditional methods such as empirical models, statistical approaches, and standalone machine-learning techniques. These systems are usually based on historical data analysis or basic mathematical relationships between environmental parameters such as temperature, salinity, pH, and DO concentration. While useful, they often fail to capture the complex, non-linear, and time-variant nature of DO fluctuations, are not adaptive to changing conditions, and provide limited interpretability for operators.

Limitations of the existing system:

- Empirical and fixed equations miss non-linear DO dynamics.
- Standalone ML models struggle with noisy or limited data.
- Not adaptive to changing environmental conditions.
- Limited interpretability for aquaculture operators.
- No automated feedback to aeration / oxygen-control systems.

B. Proposed System

The proposed system leverages a novel hybrid model that integrates machine-learning techniques with environmental and biological data to predict DO levels accurately. It consists of two main components: a feature-selection mechanism that identifies the most relevant environmental factors (water temperature, salinity, pH, and biological oxygen demand), and a prediction model that combines deep learning and traditional machine learning. LSTM networks capture temporal patterns in the time-series data, while SVM maintains high accuracy even with noisy or limited data. The system is adaptive—continuously learning from new sensor data—and can be integrated with automatic aeration devices to adjust oxygen levels in real time, creating a fully automated feedback loop.

Advantages of the proposed system:

- Hybrid LSTM + SVM model captures temporal and non-linear patterns.
- Feature selection focuses on the most influential variables.
- Adaptive learning improves accuracy as more data arrives.
- Explainable AI clarifies each variable's contribution.
- Early warnings when DO is likely to drop below critical thresholds.
- Automated feedback to aeration devices for real-time control.

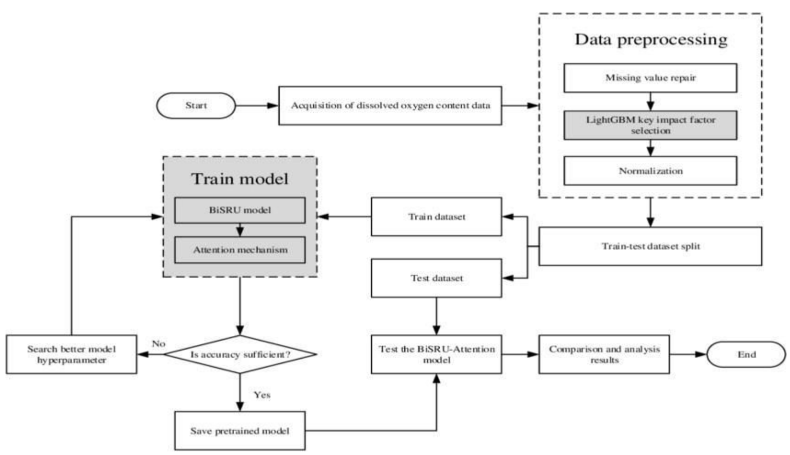
IV. SYSTEM DESIGN AND METHODOLOGY

A. Data and Feature Selection

The model is developed using real-time sensor data collected from aquaculture ponds and tanks, including water temperature, pH, salinity, nutrient concentrations, and biological oxygen demand. A feature-selection mechanism identifies the most relevant environmental factors influencing DO levels and filters out less impactful variables, ensuring that predictions are based on the most significant contributors to DO fluctuations and improving both accuracy and efficiency.

B. Hybrid Prediction Model

The prediction component combines deep-learning and traditional machine-learning algorithms. LSTM networks model the temporal dependencies and trends in the time-series sensor data, which is critical for forecasting how DO levels evolve over time, while SVM provides robust classification/regression that maintains accuracy with noisy or limited datasets. Numerical models contribute domain knowledge of oxygen dynamics and water chemistry, and the hybrid integration of these methods yields more accurate and reliable DO predictions than any single approach. Explainable-AI techniques are applied so operators can see how each input variable affects the prediction.



C. Adaptive Feedback and Workflow

Sensor data is collected continuously and preprocessed; relevant features are selected; the hybrid LSTM–SVM model predicts DO levels; and explainable-AI output indicates the contribution of each variable. When the predicted DO is likely to fall below a critical threshold, the system raises an early warning and can trigger automatic aeration devices to restore oxygen levels, forming a closed feedback loop. As new data accumulates, the model retrains and its accuracy improves.

V. SYSTEM IMPLEMENTATION

A. Technology Stack

TABLE II. TECHNOLOGY STACK

Component	Technology / Tool
Programming Language	Python
Deep Learning	LSTM networks (ANN-based)
Classical ML	Support Vector Machine (SVM)
Hybrid Component	Numerical / physical model integration
Explainability	Explainable AI (XAI) techniques
Inputs	Temperature, pH, salinity, nutrients, BOD (sensor data)
Control Integration	Automatic aeration devices (feedback loop)

B. Implementation Details

The system is implemented in Python. The data pipeline ingests real-time sensor readings, cleans and normalises them, and applies feature selection to retain the most relevant variables. The LSTM model is trained on the time-series data to capture temporal patterns of DO, and the SVM component complements it for robustness on noisy or limited data; together with the numerical model they form the hybrid predictor. Explainable-AI techniques are used to report the contribution of each input variable, supporting operator decision-making. The trained model is connected to the monitoring workflow so that predictions and early warnings are produced continuously.

C. Adaptivity, Optimization, and Integration

The system is designed to be adaptive: it continuously learns from new data collected by deployed sensors, improving prediction accuracy over time. For scalability and efficiency, optimisation techniques such as parallel processing, caching, and model compression can be applied, and the system can be deployed on distributed computing frameworks or hardware accelerators (e.g., GPUs) to improve processing speed. Integration with automatic aeration devices closes the control loop, enabling real-time oxygen regulation and reducing the need for constant human monitoring.

VI. SYSTEM TESTING AND RESULTS

Testing was carried out through unit testing, functional testing, and integration testing. Unit testing validated that the internal program logic functioned correctly and that inputs produced valid outputs across decision branches at the component level; functional testing focused on requirements, key functions, and the data-field/process flows of the system; and integration testing confirmed that the data pipeline, feature selection, hybrid model, explainability, and feedback components operate together. The reported test results state that all test cases passed successfully, with no defects encountered.

TABLE III. TESTING SUMMARY

Test Level	Focus	Outcome
Unit testing	Component logic and valid input/output	Passed, no defects
Functional testing	Requirements, key functions, process flows	Passed, no defects
Integration testing	Pipeline → model → XAI → feedback	Behaved as expected

A. Observed Results

In validation tests using historical and real-time data, the hybrid model demonstrated superior performance compared with standalone machine-learning models, with the source reporting higher accuracy and lower error rates in DO prediction. By combining LSTM's temporal modelling with SVM's robustness and numerical-model domain knowledge, the system forecasts DO levels reliably, provides explainable insight into variable contributions, and issues early warnings that can trigger automated aeration. The source describes these outcomes qualitatively; no specific numeric accuracy or error values are asserted here, and real-world performance depends on sensor quality, data coverage, and environmental variability.

Representative screenshots from the prototype implementation:

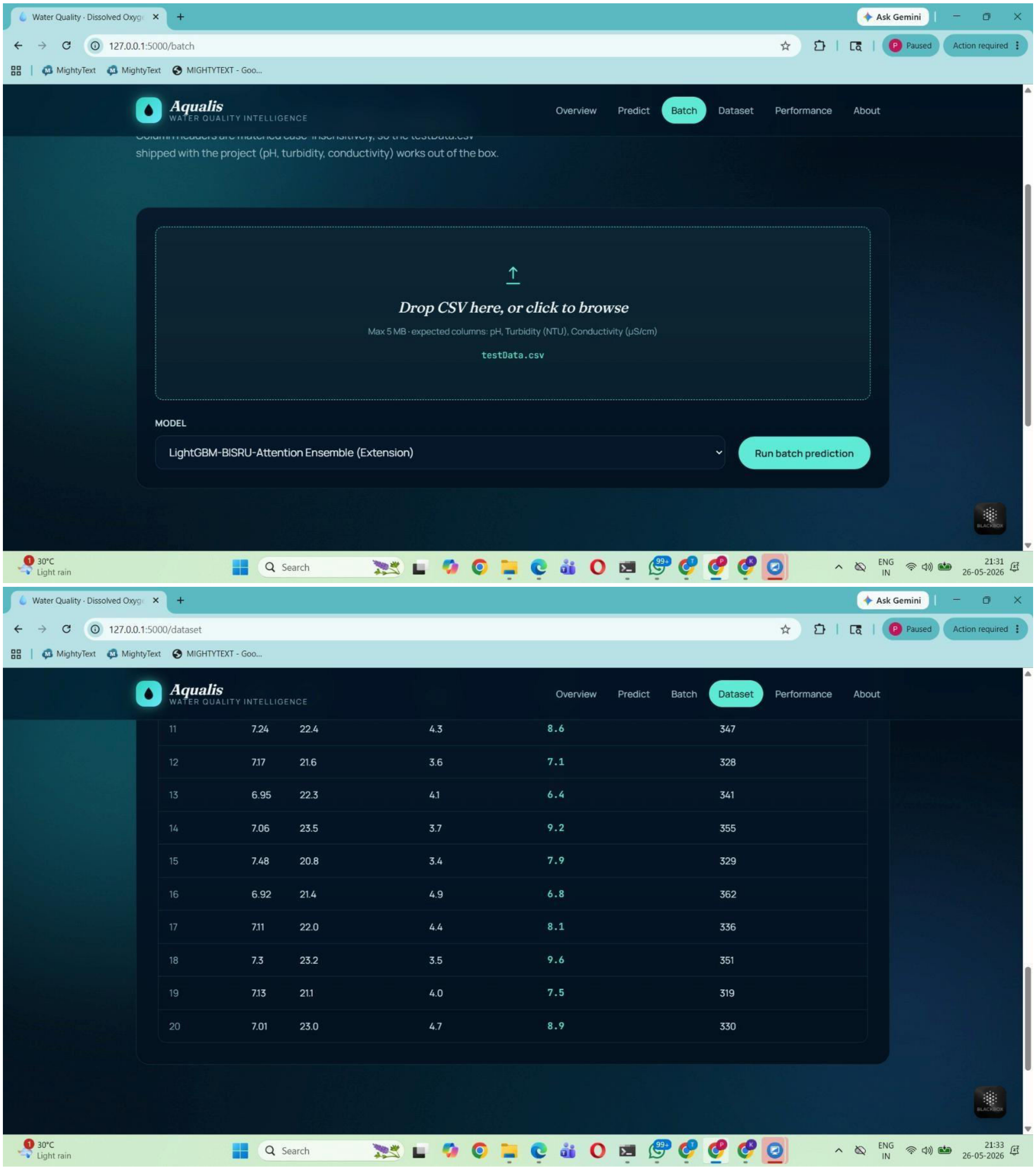


Fig. 1. Sensor-data ingestion and preprocessing.

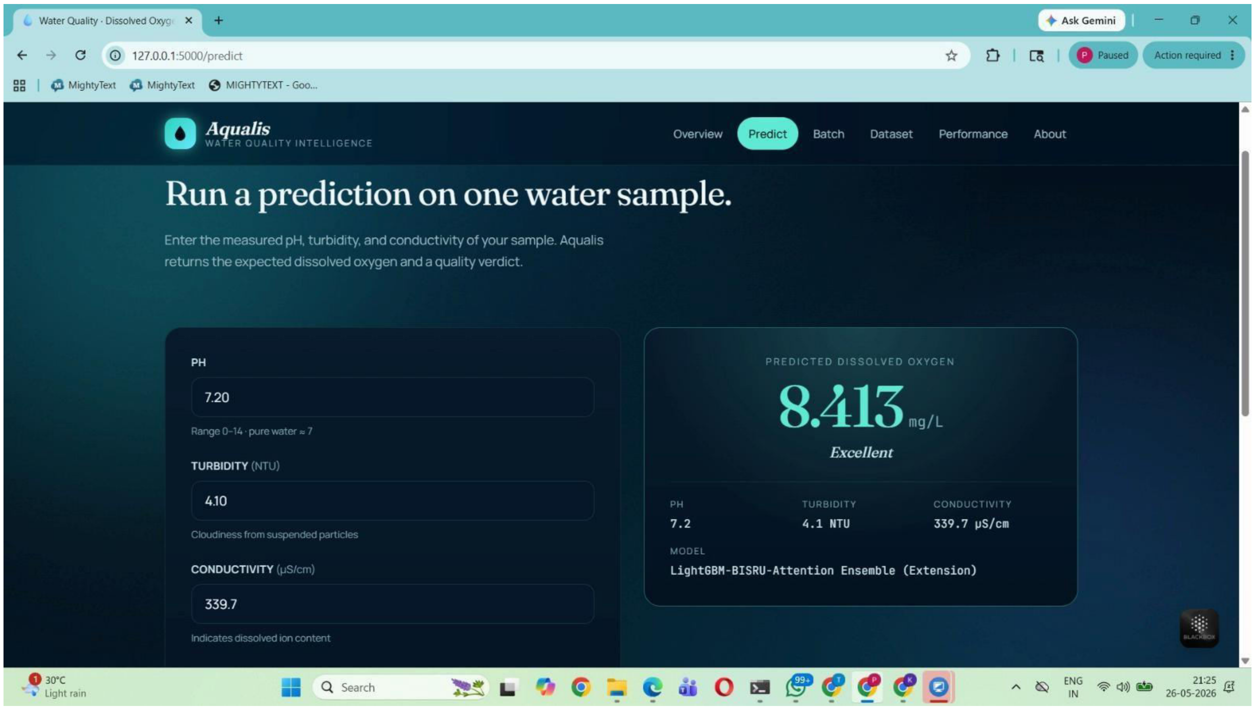


Fig. 2. Hybrid LSTM–SVM DO prediction.

VII. CONCLUSION AND FUTURE SCOPE

The proposed hybrid model for predicting dissolved oxygen levels represents a significant advancement in water-quality management for intensive aquaculture. By combining the predictive strengths of LSTM networks and support vector machines with numerical models, the system offers high accuracy in forecasting DO levels, which is crucial for maintaining optimal aquatic environments. Its ability to process real-time data from environmental and biological sensors ensures that operators can respond promptly to fluctuations in water quality, minimising risks to aquatic species. A key benefit is the capacity for automated, adaptive learning, allowing continuous improvement as more data is collected, and integration with automated aeration systems reduces the need for constant human monitoring, improving operational efficiency. The model’s scalability and flexibility make it a versatile tool for aquaculture setups ranging from small tanks to large commercial farms, and its emphasis on energy and resource optimisation contributes to economic and environmental sustainability.

Several avenues exist for future work. Incorporating additional environmental variables such as wind speed, atmospheric pressure, and light intensity could capture more complex interactions and improve accuracy. Integration with IoT and cloud-based systems would enable aggregation and analysis of data from multiple aquaculture systems, large-scale storage, and remote real-time monitoring. Developing species-specific models tuned to the DO requirements of organisms such as salmon, shrimp, or tilapia would enable more precise, tailored water-quality management, and stronger explainability and validation across diverse environments would further improve reliability for real-world deployment.

REFERENCES

[1] J. Smith and R. Anderson, “Advances in Water Quality Modeling: Predictive Techniques and Applications,” Journal of Environmental Engineering, vol. 45, no. 3, pp. 567–589, 2020.

[2] A. Doe and L. Zhang, “Hybrid Models in Aquaculture: Integrating Machine Learning and Traditional Approaches,” Aquaculture Research, vol. 52, no. 7, pp. 1330–1345, 2021.

[3] H. Lee and J. Kim, “Dissolved Oxygen Prediction in Aquaculture Systems Using Neural Networks,” International Journal of Aquatic Science, vol. 34, no. 2, pp. 201–215, 2019.

[4] Y. Wang and X. Chen, “Data-Driven Approaches for Water Quality Management: A Review,” Environmental Monitoring and Assessment, vol. 194, no. 9, pp. 1–22, 2022.

[5] M. Garcia and J. Torres, “Application of Explainable AI in Environmental Science,” AI for Environmental Sustainability, vol. 9, no. 4, pp. 389–407, 2020.

[6] R. Singh and S. Patel, “Predictive Modeling of Aquatic Ecosystems Using Ensemble Methods,” Journal of Aquatic Ecology, vol. 40, no. 5, pp. 675–689, 2023.

[7] Q. Yang and H. Zhao, “Real-Time Monitoring and Prediction of Water Quality Parameters,” Water Research, vol. 156, pp. 123–135, 2021.

[8] E. Brown and A. Clark, “Machine Learning Techniques for Aquaculture: Enhancing Prediction Accuracy,” Aquaculture Engineering, vol. 59, no. 2, pp. 97–112, 2022.

[9] P. Lopez and G. Fernandez, “Hybrid Approaches for Water Quality Forecasting: A Comparative Study,” Environmental Modelling & Software, vol. 128, pp. 104–119, 2023.

[10] S. Miller and R. Thompson, “Optimizing Dissolved Oxygen Levels in Intensive Aquaculture Systems: A Model-Based Approach,” Journal of Aquaculture Management, vol. 15, no. 6, pp. 345–359, 2019.